

国内社会科学系データアーカイブの横断検索に向けての試行と検討

国立情報学研究所 ○船守美穂, 林正治, 加藤文彦 東京大学 三輪哲
立教大学 朝岡誠, 高橋かおり 関西学院大学 前田豊

1 問題の所在と解決の方向性

国内には複数の、社会科学系のデータを収集・管理し、求めに応じて提供する社会科学系データアーカイブ（以下、SSDA）が存在する。いずれの SSDA も、社会を洞察するのに貴重なデータを扱っていながらも、一箇所で集約的にデータを検索することはできず、データ利用者はそれぞれの SSDA において、自身が必要とするデータセットの存在の有無を調べる必要があり、極めて非効率であった。

今回は、国立情報学研究所（NII）が全国に対して提供する学術情報流通のためのサービスを利用して横断型 SSDA を構築する際の、1）既存システムを組み合わせる方法等、アーキテクチャ面と、2）メタデータ等、DB の構造・運用面について、検討を行った。

2 横断型 SSDA のアーキテクチャ面の検討

NII では現在、論文等の学術資料を保存・公開する「機関リポジトリのクラウドサービス（JAIRO Cloud）」と、学術資料を検索できる「NII 学術情報ナビゲータ（CiNii）」とを提供している。これらは世界的なオープンサイエンスに向けての流れを受け、2020 年度以降、研究データに対しても利用可能となる予定である。この枠組みを利用した横断型 SSDA の構築を、I）SSDA データセットを登録する機関リポジトリ（IR）、II）登録されたデータセットを検索する枠組みの2側面から検討した。

既存の SSDA データセットを登録する IR としては、A) 横断型 SSDA 専用の IR を一つ立て、そこに全てのデータセットを集約する方法と、B) それぞれの SSDA が運営される大学の IR にデータセットを登録する方法とがある。A) は分かりやすい一方、各 SSDA の担当者、あるいは各 SSDA の存在そのものが経年的に変化することを踏まえると、横断型 SSDA の管理主体なしでは運営が困難となることが予想される。B) は、各大学の IR 担当者にデータセットの永続的管理を任せられるという点で優れているが、国内全ての大学が IR を有するわけではなく、また SSDA のデータセットを各大学の IR に統合するかどうかは、各大学の方針に依存する。A) B)、あるいは両者を併用するかは、各 SSDA および横断型 SSDA の管理運営主体を見据えながら、検討していく必要がある。

検索の枠組みとしては、a) SSDA データセットが登録された IR 内の検索機能、b) IR 横断的な検索サービス（JAIRO）、c) IR とその他の学術情報 DB を横断的に検索する CiNii を利用する方法とがある。段階的に検索の範囲が拡大し、また検索項目が粗くなる。どの検索枠組みを利用するかは、利用として、社会科学限定的、学際領域型研究、一般利用のいずれを想定するかにより変わる。

3 複数 SSDA を統合する際のデータベース構造・運用面の検討

今回は JAIRO Cloud に IR のインスタンスを一つ立て、横断型 SSDA を仮構築し（Aa 方式）、複数 SSDA を統合する際の、DB 構造・運用面の課題の洗い出しを行った。①東京大学社会科学研究所の運営する SSJDA と、②立教大学社会調査データアーカイブ（RUDA）からそれぞれ、3 と 10 のデータセットの登録を得た。メタデータの規格としては、SSDA の国際規格である DDI を用いた。

独自のメタデータを利用していた両 SSDA とも、この規格にメタデータをマッピングすることはそれほど問題なかったが、各フィールドに入力する内容の統一性および、それを誰が担保するかなど、運用面において課題があることが判明した。SSDA ごとに独自の入力ルールがあり（ex. 「KW」が自由記述又は選択方式）、また単一 SSDA 内でもデータセットにより入力方針が異なり、これを統一しないと意味のある横断検索は実現しないが、そのためには相当の時間と労力を要する。この統一性の確保は、参加する SSDA やデータセットが拡大するとともに難しくなる。また「科研費番号」など、揺らぎのない項目でも、これが正確に入力されるためには、SSDA 管理者が必要と認識された。